Optimizing Balance and Sample Size in Matching Methods for Causal Inference[1]

Gary King[2]

Institute for Quantitative Social Science
Harvard University

(Talk at UCLA, 3/1/2013)

[1] Joint work with Christopher Lucas and Richard Nielsen
[2] GaryKing.org.

Overview

Overview

- Problem: Model dependence (review)

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)
- Solution: A New Approach — Easier & More Powerful

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the "n-imbalance frontier"

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix *n* and don't guarantee imbalance
  - Others fix imbalance and don't guarantee *n*
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the "*n*-imbalance frontier"
  - Imbalance metric choice defines the frontier

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the "n-imbalance frontier"
  - Imbalance metric choice defines the frontier
- Side point:

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the "n-imbalance frontier"
  - Imbalance metric choice defines the frontier
- Side point:
  - Problem: A commonly used method increases imbalance!

Overview

- Problem: Model dependence (review)
- Solution: Matching to reduce model dependence (review)
- Problem: Matching methods optimize only 1 of 2 parameters
  - Some fix n and don't guarantee imbalance
  - Others fix imbalance and don't guarantee n
  - (Plus: Matching methods optimize a different "imbalance" thanks post-hoc checks)
- Solution: A New Approach — Easier & More Powerful
  - We estimate the "n-imbalance frontier"
  - Imbalance metric choice defines the frontier
- Side point:
  - Problem: A commonly used method increases imbalance!
  - Solution: Other methods & our approach do not share this problem

Model Dependence Example

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

- Data: 124 Post-World War II civil wars

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,. . .

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,...
- Counterfactual question: Switch UN intervention for each war

Model Dependence Example

Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,. . .
- Counterfactual question: Switch UN intervention for each war
- Data analysis: Logit model

Model Dependence Example

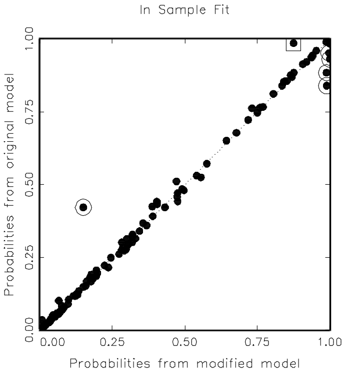
Replication: Doyle and Sambanis, APSR 2000
(King and Zeng, 2007)

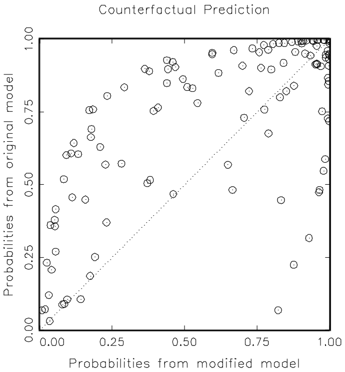
- Data: 124 Post-World War II civil wars
- Dependent var: peacebuilding success
- Treatment: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status,...
- Counterfactual question: Switch UN intervention for each war
- Data analysis: Logit model
- The question: How *model dependent* are the results?

Two Logit Models, Apparently Similar Results

	Original "Interactive" Model			Modified Model		
Variables	Coeff	SE	P-val	Coeff	SE	P-val
Wartype	−1.742	.609	.004	−1.666	.606	.006
Logdead	−.445	.126	.000	−.437	.125	.000
Wardur	.006	.006	.258	.006	.006	.342
Factnum	−1.259	.703	.073	−1.045	.899	.245
Factnum2	.062	.065	.346	.032	.104	.756
Trnsfcap	.004	.002	.010	.004	.002	.017
Develop	.001	.000	.065	.001	.000	.068
Exp	−6.016	3.071	.050	−6.215	3.065	.043
Decade	−.299	.169	.077	−0.284	.169	.093
Treaty	2.124	.821	.010	2.126	.802	.008
UNOP4	3.135	1.091	.004	.262	1.392	.851
Wardur*UNOP4	—	—	—	.037	.011	.001
Constant	8.609	2.157	0.000	7.978	2.350	.000
N		122			122	
Log-likelihood		-45.649			-44.902	
Pseudo R^2		.423			.433	

Doyle and Sambanis: Model Dependence





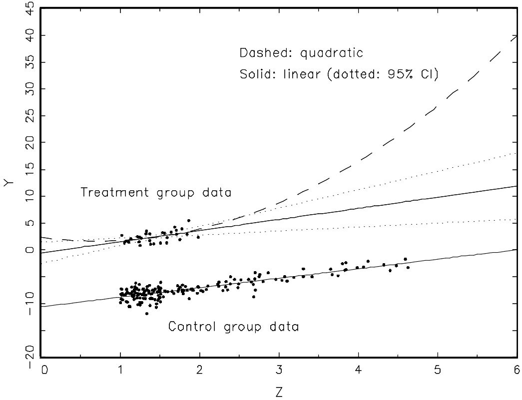
Model Dependence: A Simpler Example

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

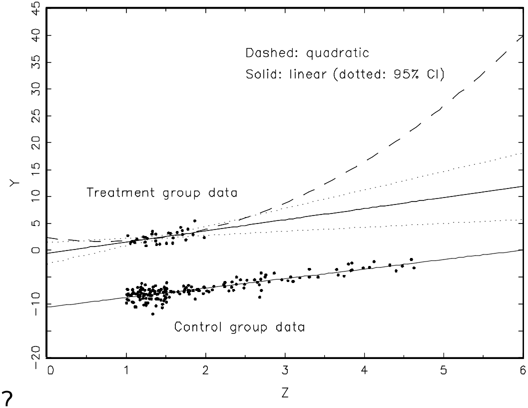
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)



Model Dependence: A Simpler Example

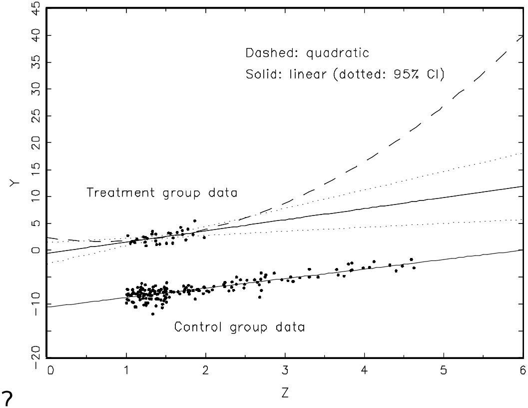
(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

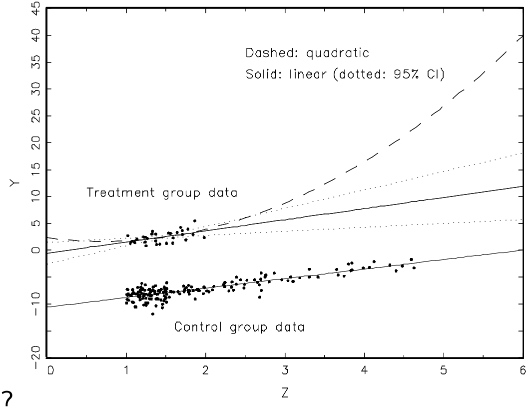


What to do?

- Preprocess I: Eliminate extrapolation region

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

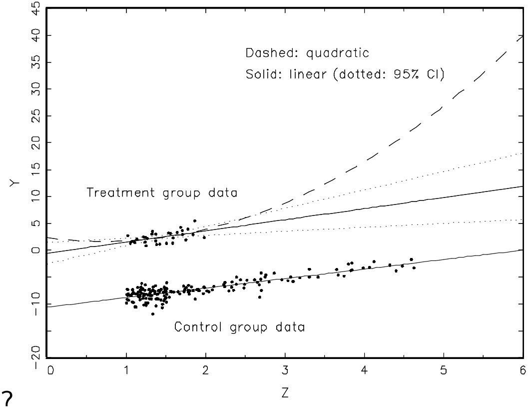


What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance

Matching within the Interpolation Region

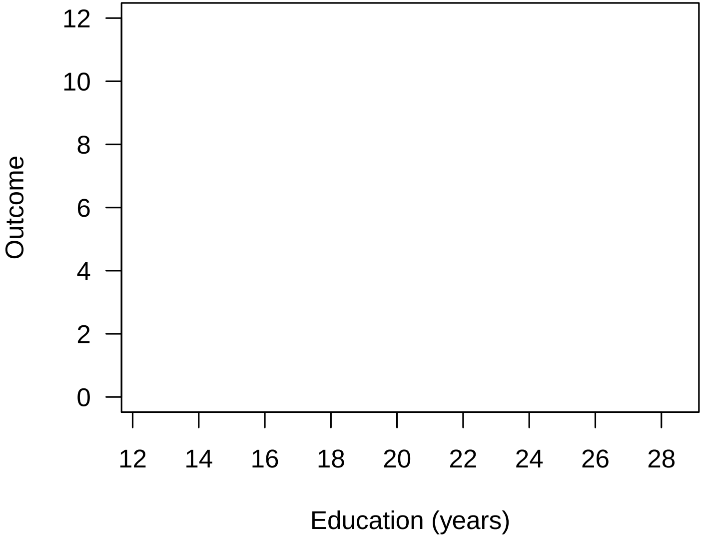
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

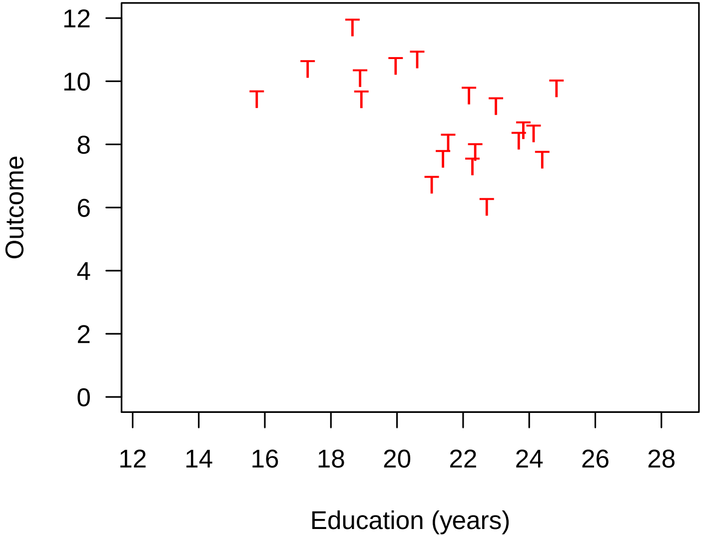
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



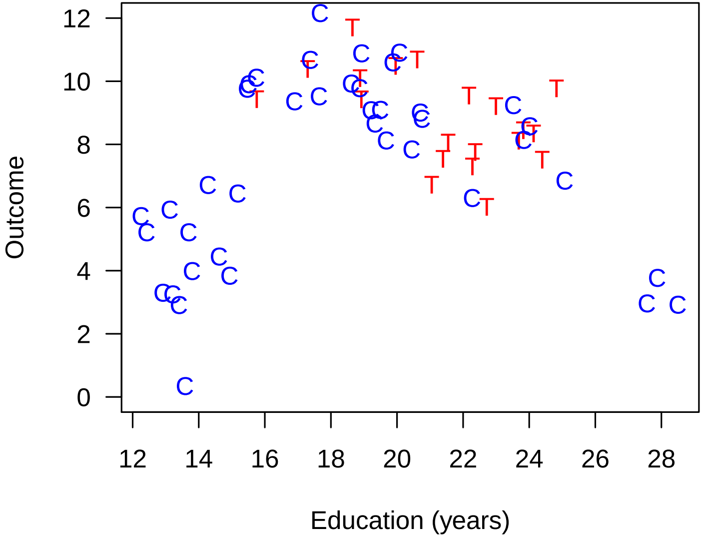
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



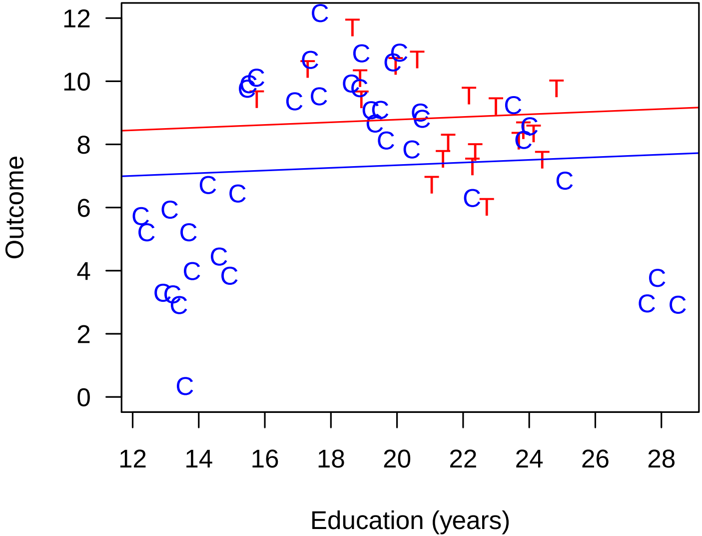
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



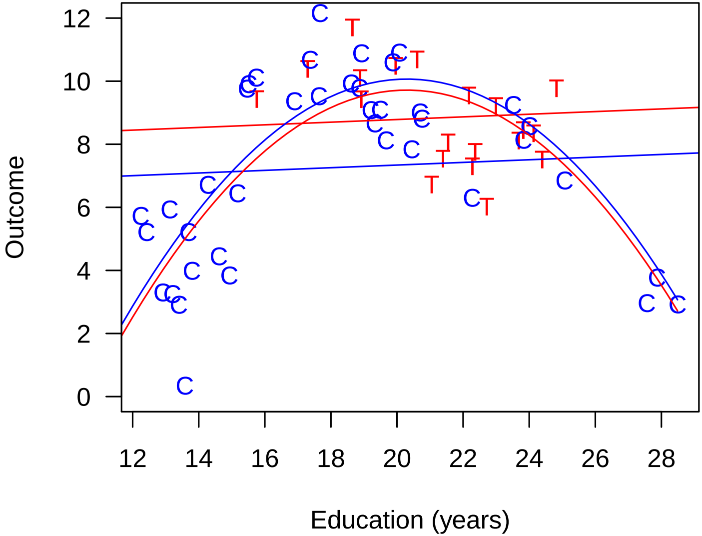
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



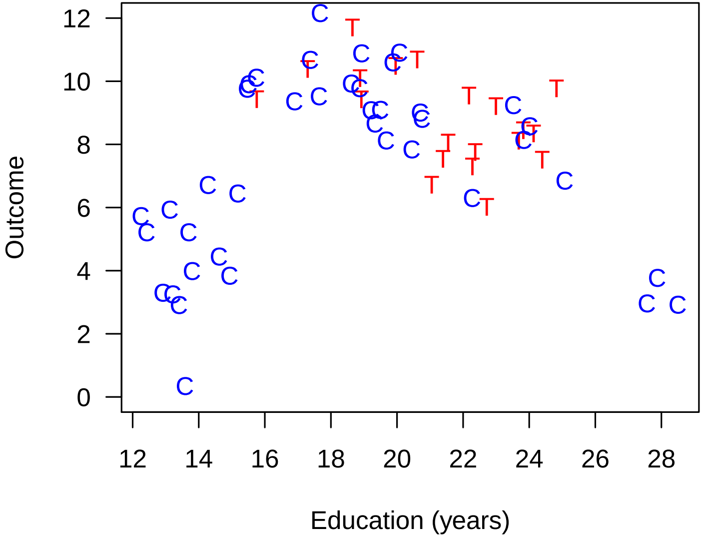
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



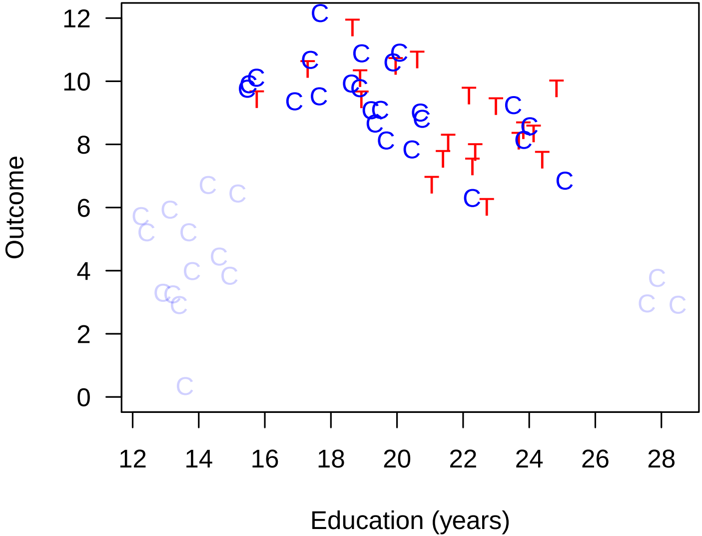
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



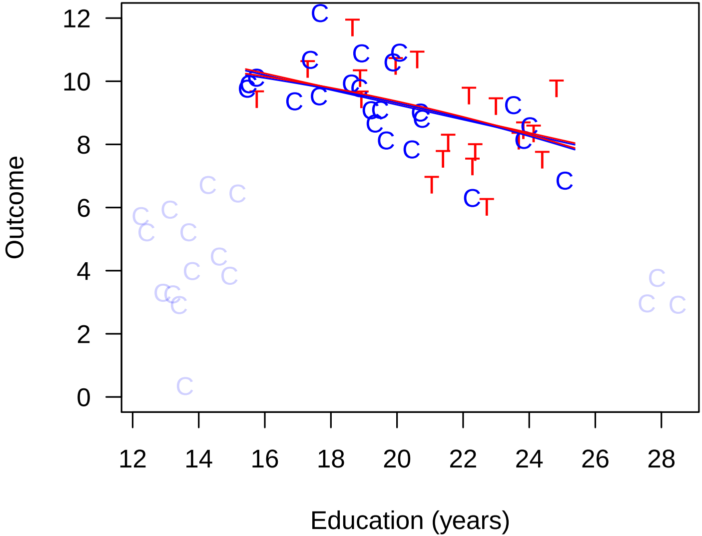
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

Matching reduces model dependence, bias, and variance

How Matching Works

How Matching Works

- Notation:

How Matching Works

- Notation:

  Y_i Dependent variable

How Matching Works

- Notation:

  Y_i Dependent variable

  T_i Treatment variable (0/1, or more general)

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates

How Matching Works

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1, or more general)
  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

How Matching Works

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1, or more general)
  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

How Matching Works

- Notation:

  Y_i Dependent variable

  T_i Treatment variable (0/1, or more general)

  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - {\color{red} Y_i(T_i = 0)} \\ &= \text{observed} \quad - {\color{red} \textit{unobserved}} \end{aligned}$$

How Matching Works

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1, or more general)
  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)

How Matching Works

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1, or more general)
  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
- Prune unmatched units to improve balance (so X is unimportant)

How Matching Works

- Notation:

  Y_i Dependent variable

  T_i Treatment variable (0/1, or more general)

  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
- Prune unmatched units to improve balance (so X is unimportant)
- QoI: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \text{mean}_{i \in \{T_i = 1\}}(\text{TE}_i)$$

How Matching Works

- Notation:

  Y_i Dependent variable
  T_i Treatment variable (0/1, or more general)
  X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
- Prune unmatched units to improve balance (so X is unimportant)
- QoI: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \text{mean}_{i \in \{T_i = 1\}}(\text{TE}_i)$$

- or Feasible Average Treatment effect on the Treated: FSATT

Method 1: Mahalanobis Distance Matching

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - $\text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - $\text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .

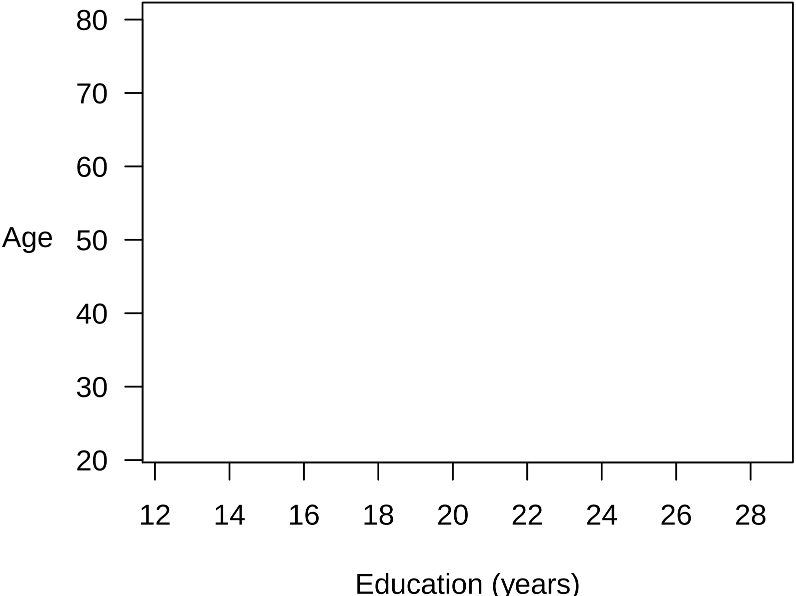
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

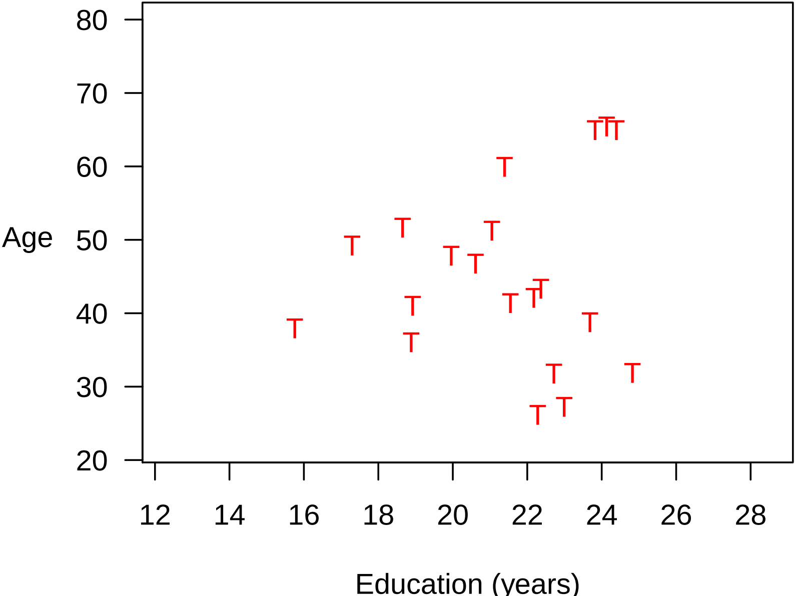
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - $\text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance>*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

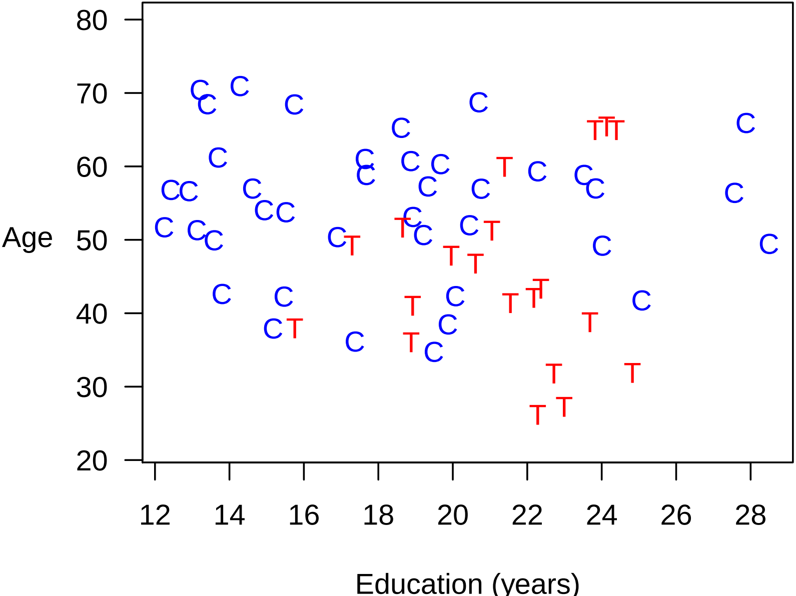
Mahalanobis Distance Matching



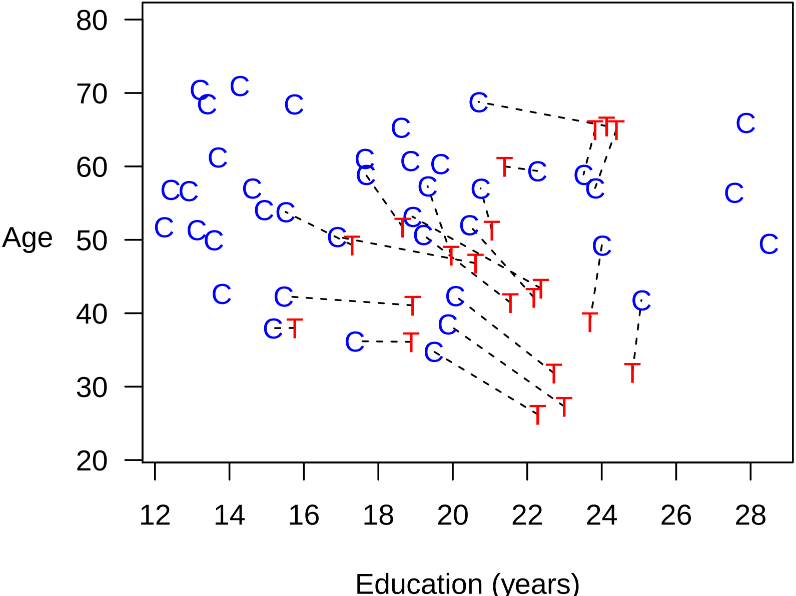
Mahalanobis Distance Matching



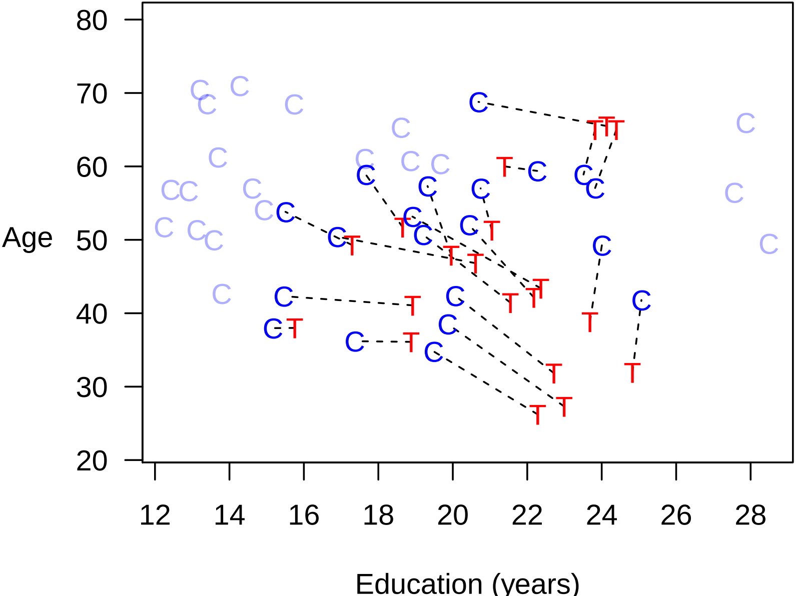
Mahalanobis Distance Matching



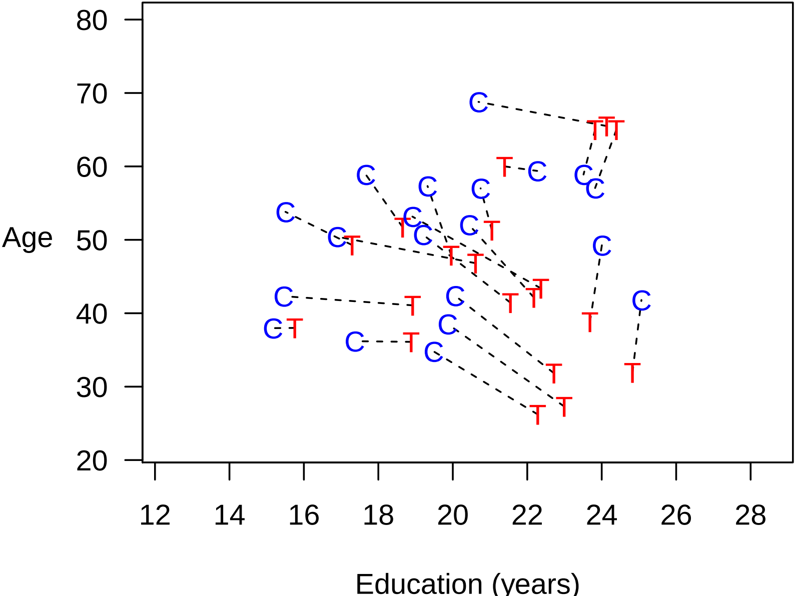
Mahalanobis Distance Matching



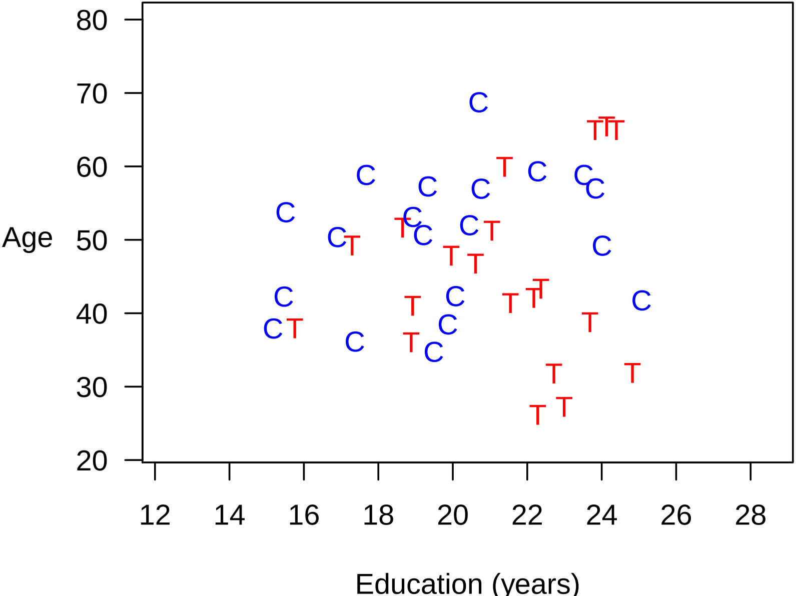
Mahalanobis Distance Matching



Mahalanobis Distance Matching



Mahalanobis Distance Matching



Method 2: Propensity Score Matching

Method 2: Propensity Score Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar
     $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
    - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

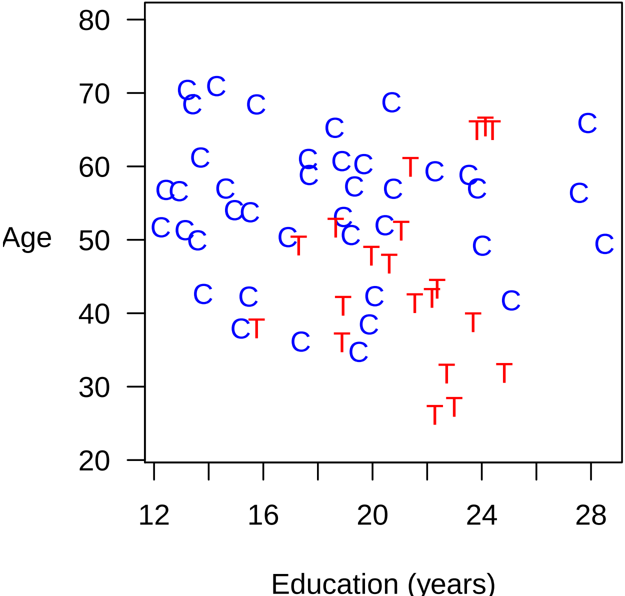
Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance>*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

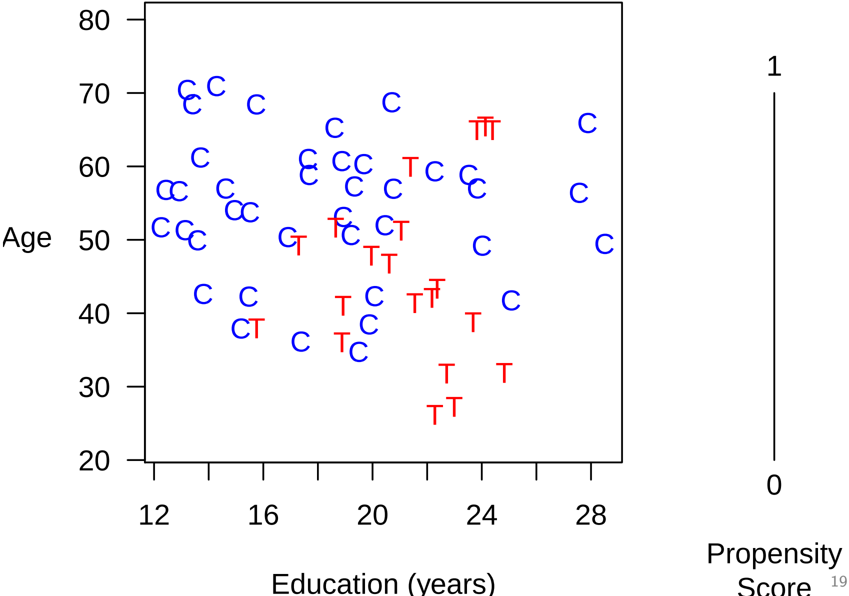
Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance>*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .

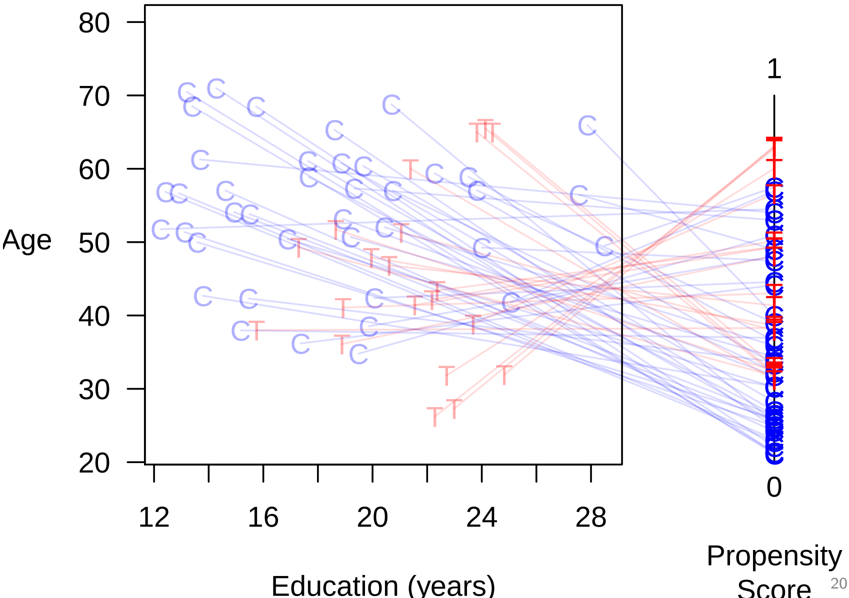
Propensity Score Matching



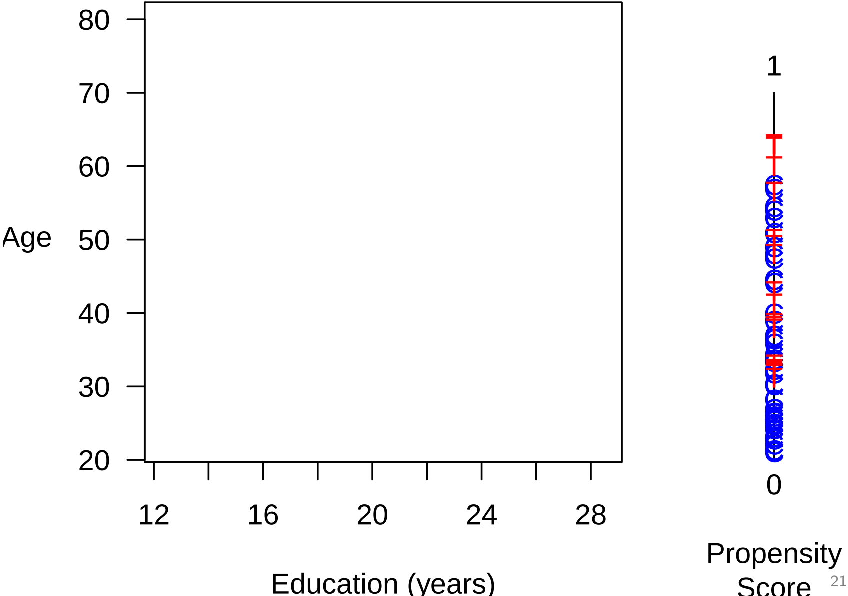
Propensity Score Matching



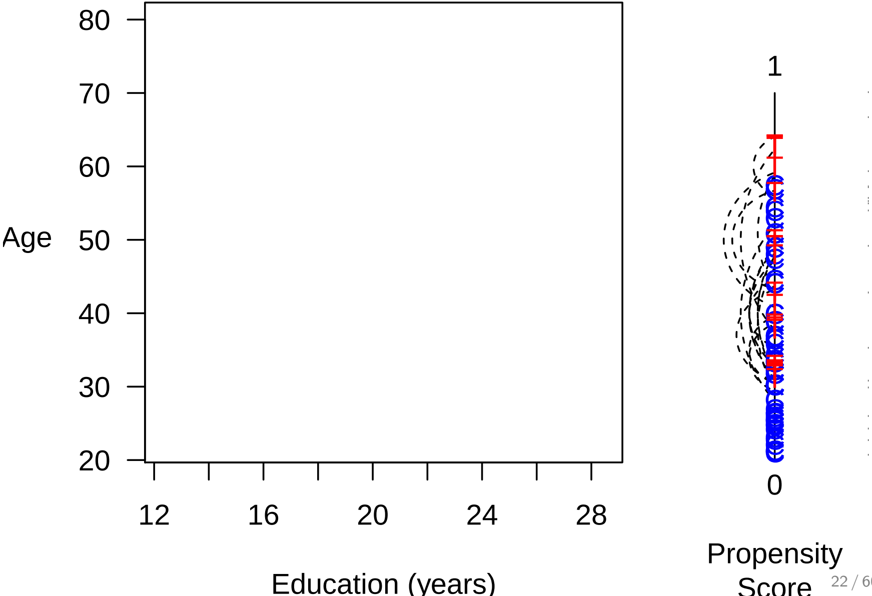
Propensity Score Matching



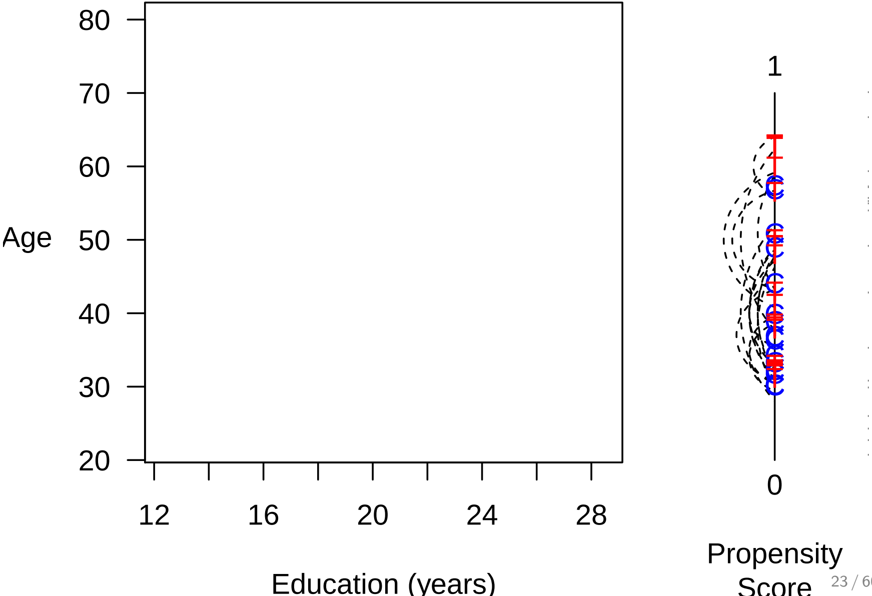
Propensity Score Matching



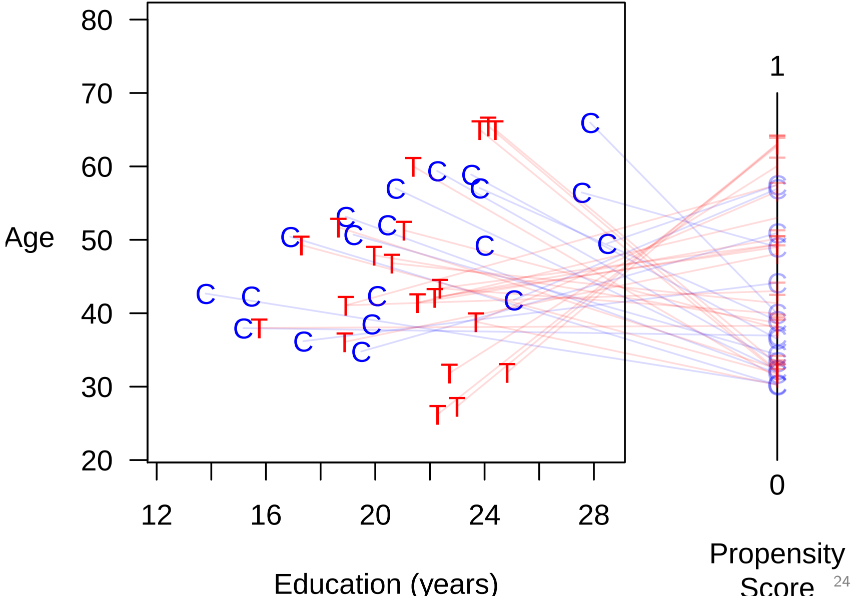
Propensity Score Matching



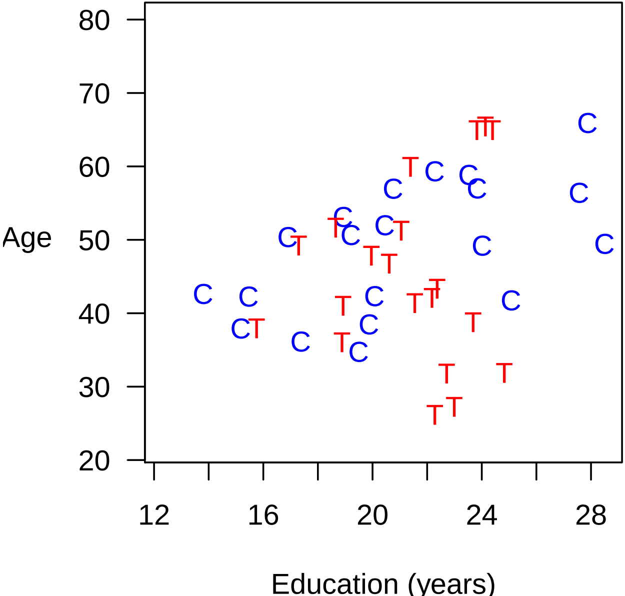
Propensity Score Matching



Propensity Score Matching



Propensity Score Matching



Method 3: Coarsened Exact Matching

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, . . .

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
    - Temporarily coarsen X as much as you're willing
        - e.g., Education (grade school, high school, college, graduate)
2. Estimation Difference in means or a model
3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
    - Temporarily coarsen X as much as you're willing
        - e.g., Education (grade school, high school, college, graduate)
        - Easy to understand, or can be automated as for a histogram
    - Apply exact matching to the coarsened X, $C(X)$

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
    - Temporarily coarsen X as much as you're willing
        - e.g., Education (grade school, high school, college, graduate)
        - Easy to understand, or can be automated as for a histogram
    - Apply exact matching to the coarsened X, $C(X)$
        - Sort observations into strata, each with unique values of $C(X)$

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
2. Estimation Difference in means or a model
3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
    - Temporarily coarsen X as much as you're willing
        - e.g., Education (grade school, high school, college, graduate)
        - Easy to understand, or can be automated as for a histogram
    - Apply exact matching to the coarsened X, $C(X)$
        - Sort observations into strata, each with unique values of $C(X)$
        - Prune any stratum with 0 treated or 0 control units
    - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds
3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

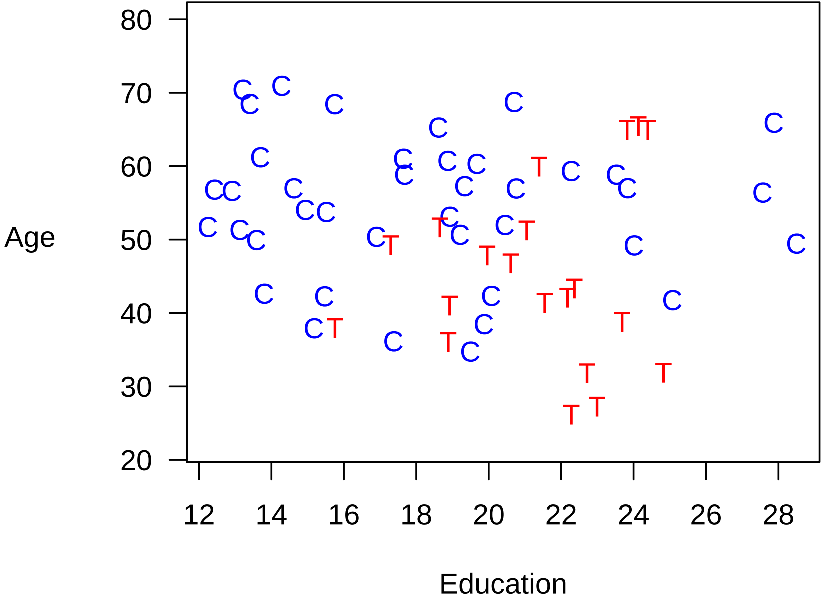
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds
   - Can apply other matching methods within CEM strata (inherit CEM's properties)
3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

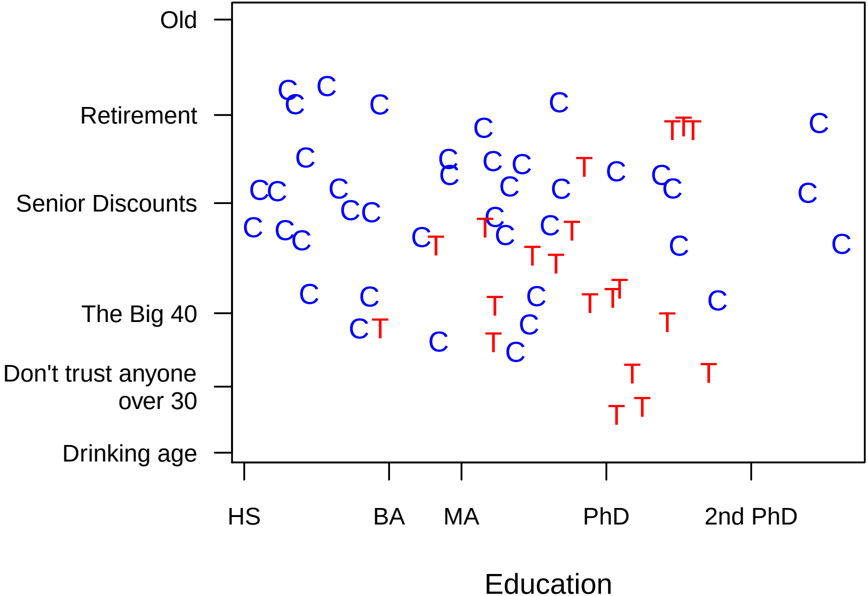
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds
   - Can apply other matching methods within CEM strata (inherit CEM's properties)
3. Checking Determine matched sample size, tweak, repeat, ...
   - Easier, but still iterative

Coarsened Exact Matching

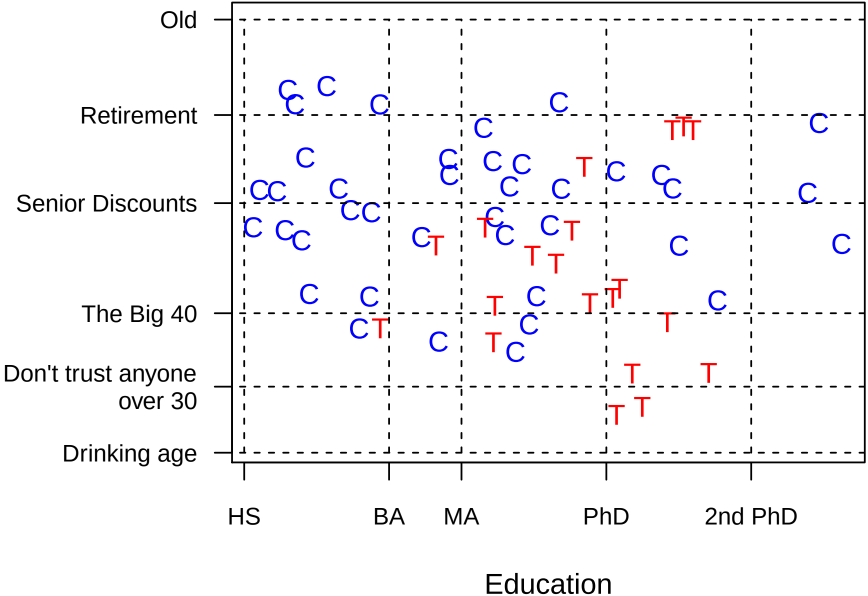
Coarsened Exact Matching



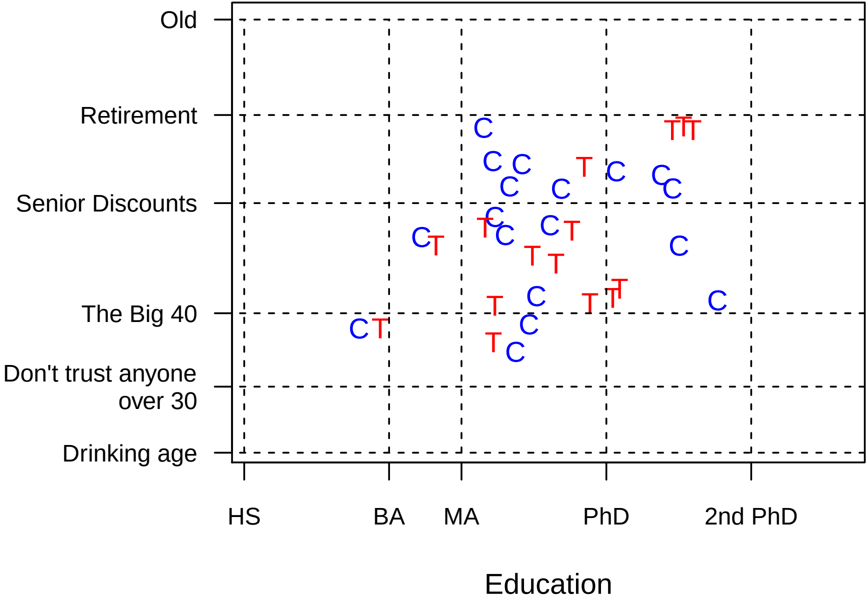
Coarsened Exact Matching



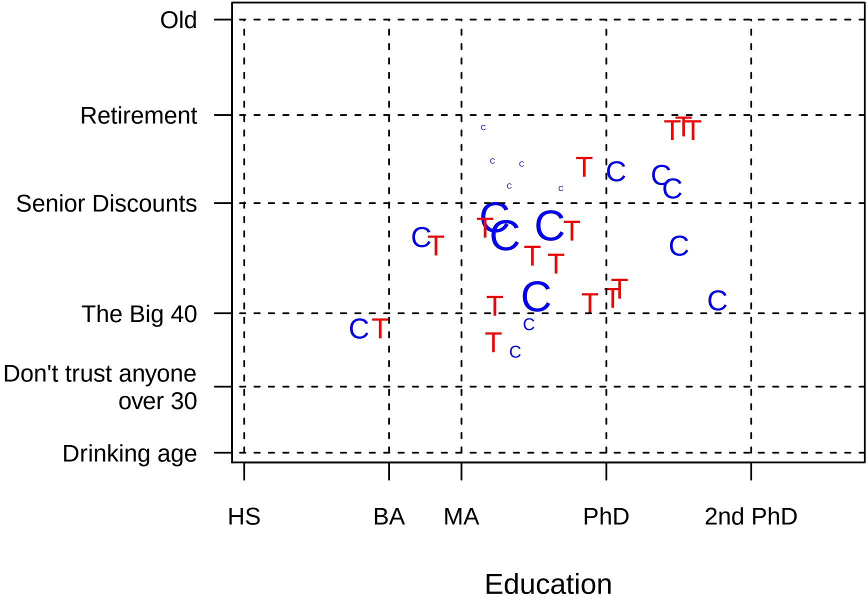
Coarsened Exact Matching



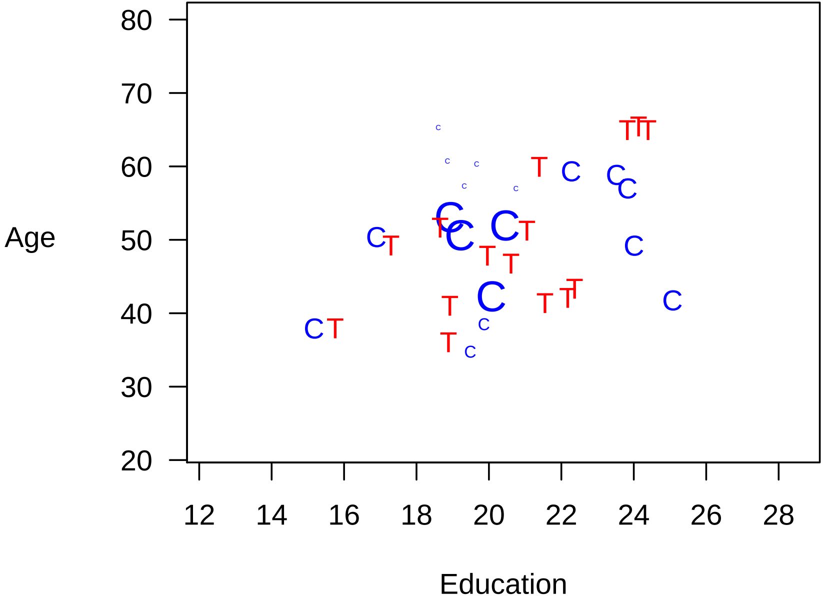
Coarsened Exact Matching



Coarsened Exact Matching



Coarsened Exact Matching



The Bias-Variance Trade Off in Matching

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f(\text{imbalance}\ldots)$
  $\rightsquigarrow$ we measure imbalance instead

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f($imbalance$\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f($matched sample size$\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f(\text{imbalance}\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f(\text{matched sample size}\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f(\text{imbalance}\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f(\text{matched sample size}\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Measuring Imbalance

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f($imbalance$\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f($matched sample size$\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f(\text{imbalance}\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f(\text{matched sample size}\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f(\text{imbalance}\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f(\text{matched sample size}\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f,g;H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f(\text{imbalance}\ldots)$
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f(\text{matched sample size}\ldots)$
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

  - Difference of multivariate histograms (L2)

The Matching Frontier

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

  - Difference of multivariate histograms (L2)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

  - Difference of multivariate histograms (L2)
- The metric defines the "n-imbalance frontier" (lowest imbalance for each n)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

  - Difference of multivariate histograms (L2)
- The metric defines the "n-imbalance frontier" (lowest imbalance for each n)
- Choose a matching solution (trading off bias and variance)

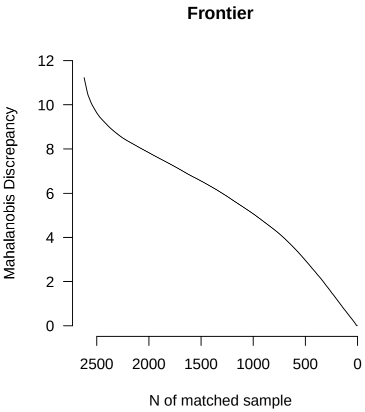
The Matching Frontier

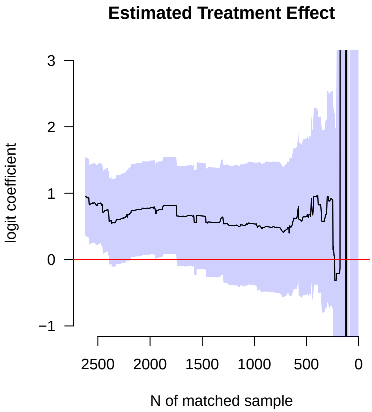
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
- Choose an imbalance metric
  - Classic measure: Difference of means (for each variable)
  - Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
  - Difference of multivariate histograms (L1):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

  - Difference of multivariate histograms (L2)
- The metric defines the "n-imbalance frontier" (lowest imbalance for each n)
- Choose a matching solution (trading off bias and variance)
- Result: Optimal. No need to iterate. Choice of solution left to researcher.

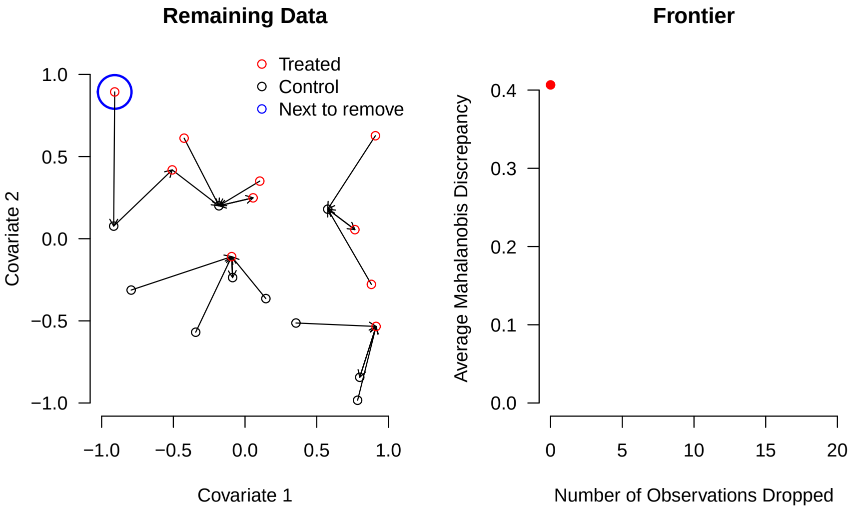
Example Frontier, and Results



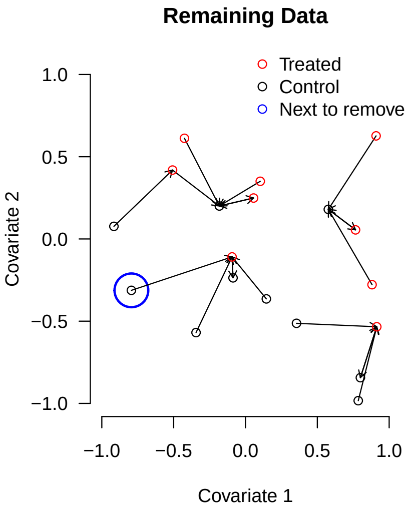


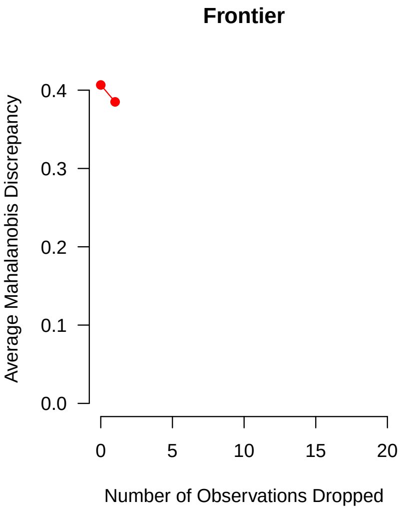
Constructing the Mahalanobis Frontier

Constructing the Mahalanobis Frontier

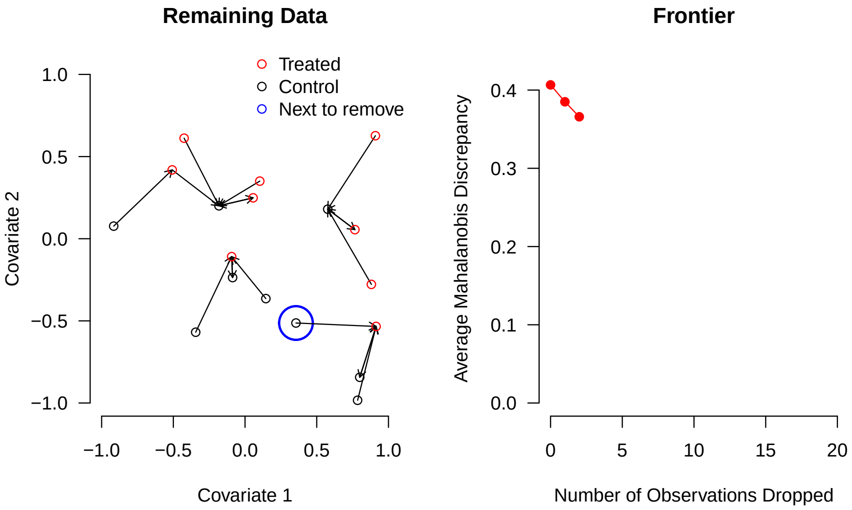


Constructing the Mahalanobis Frontier

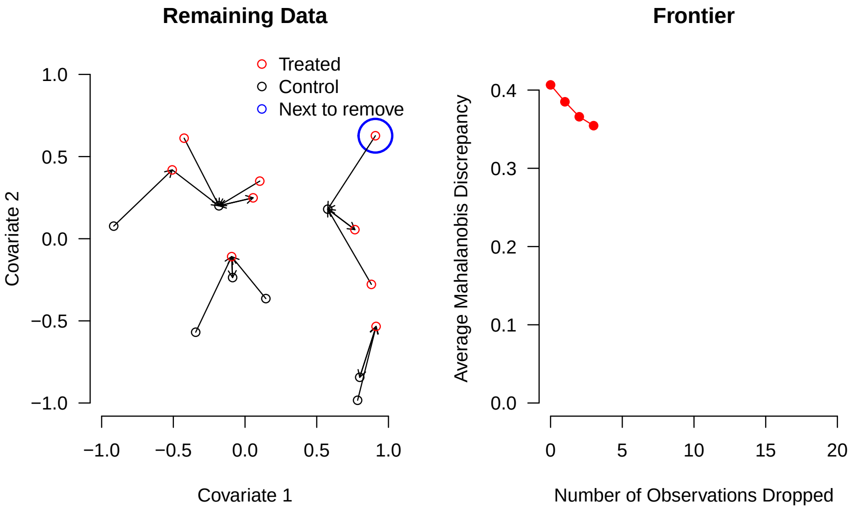




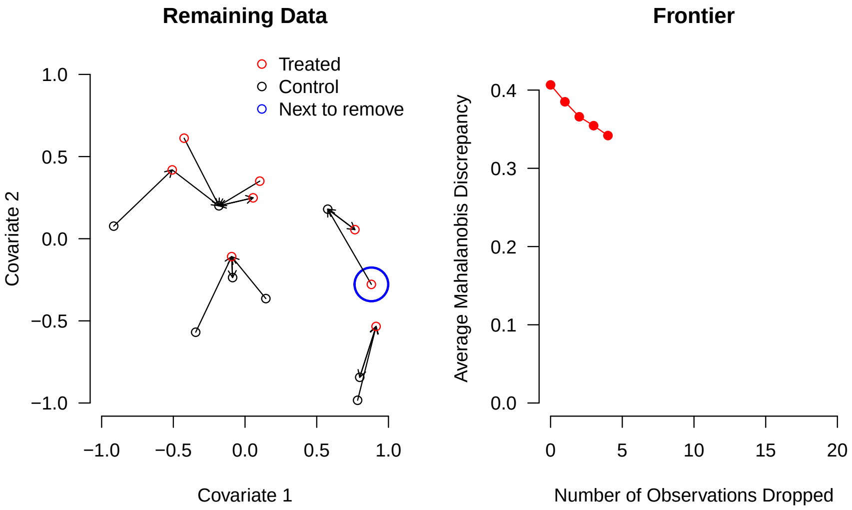
Constructing the Mahalanobis Frontier



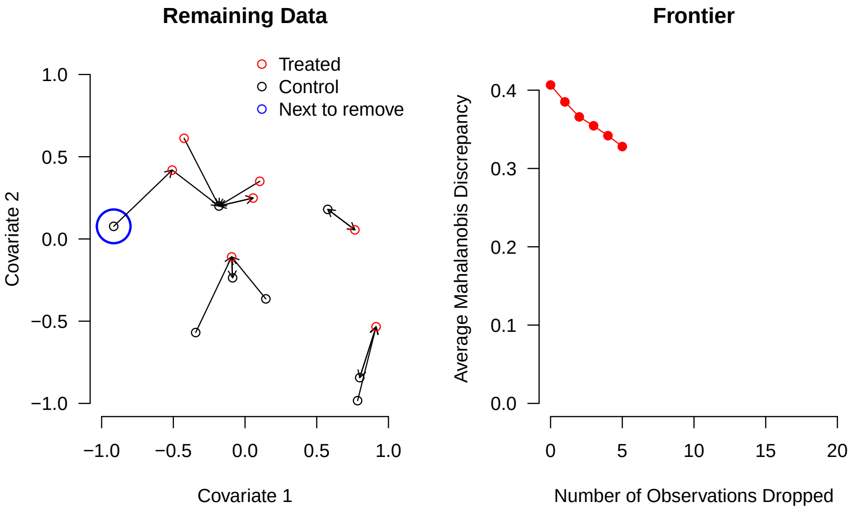
Constructing the Mahalanobis Frontier



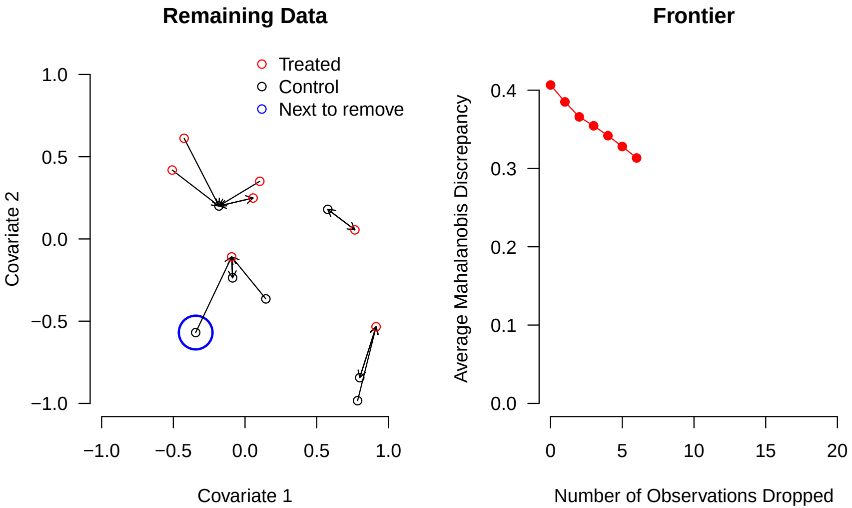
Constructing the Mahalanobis Frontier



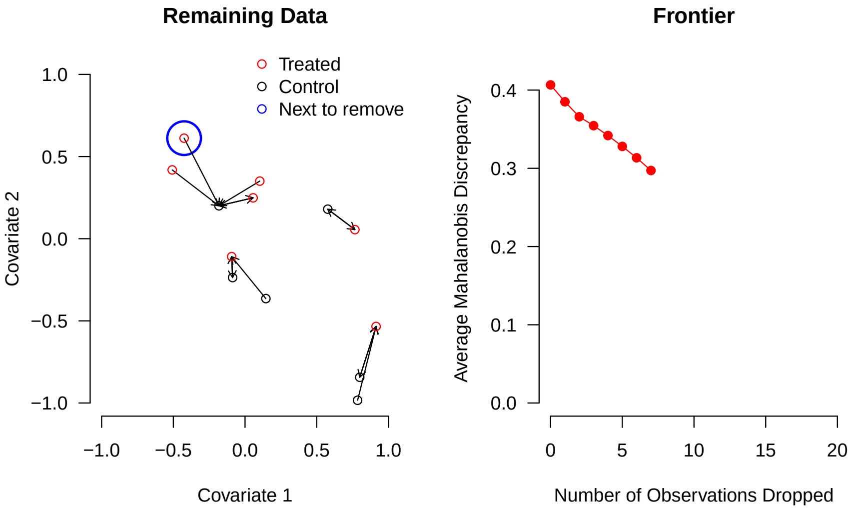
Constructing the Mahalanobis Frontier



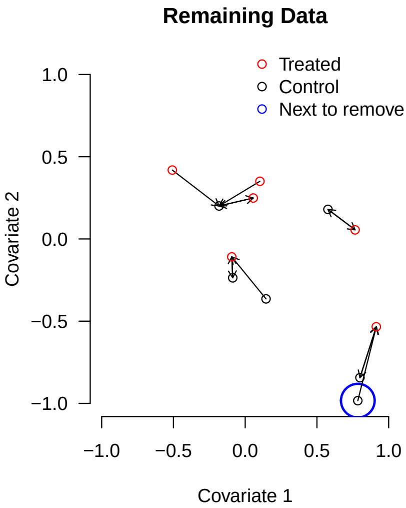
Constructing the Mahalanobis Frontier

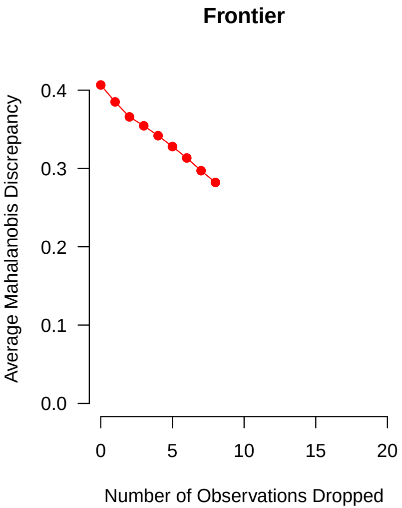


Constructing the Mahalanobis Frontier

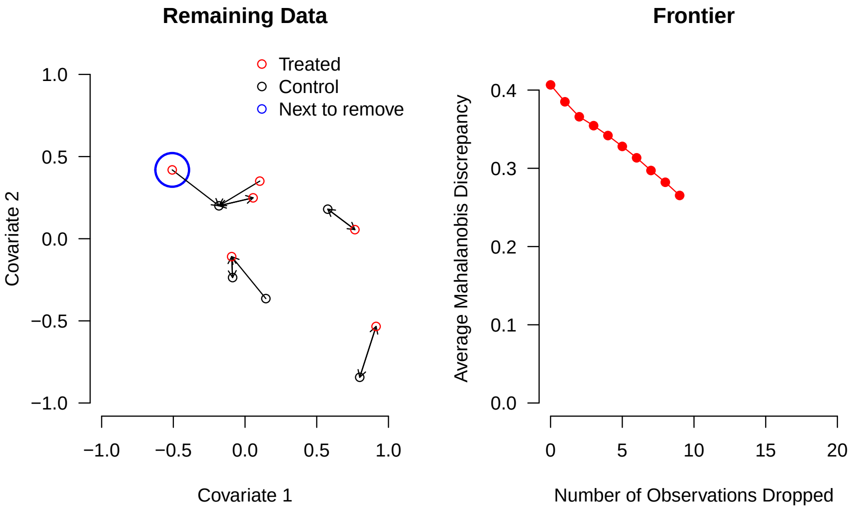


Constructing the Mahalanobis Frontier

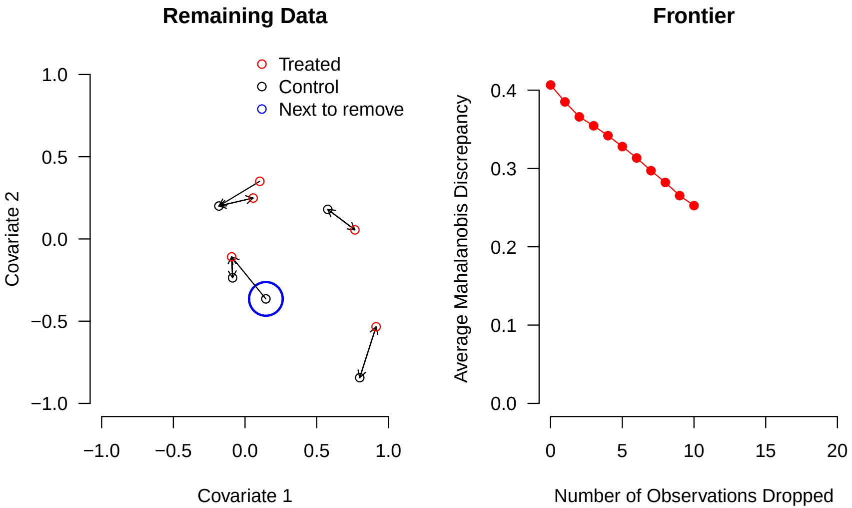




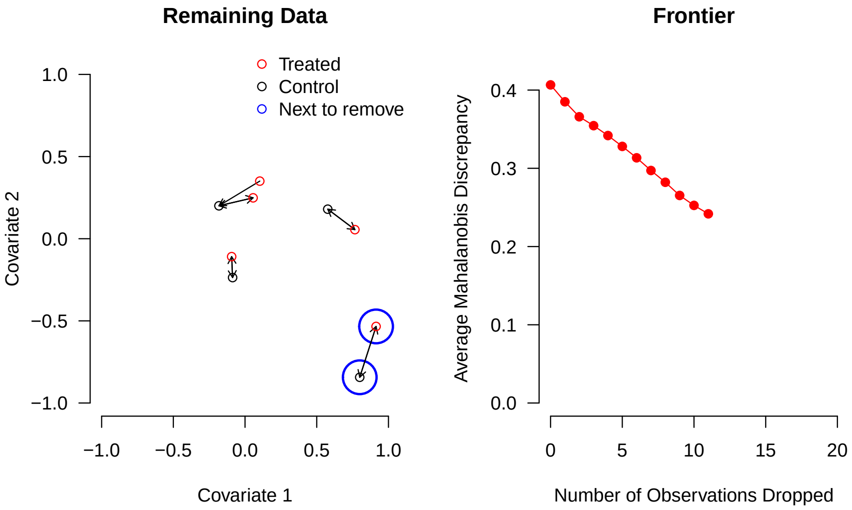
Constructing the Mahalanobis Frontier



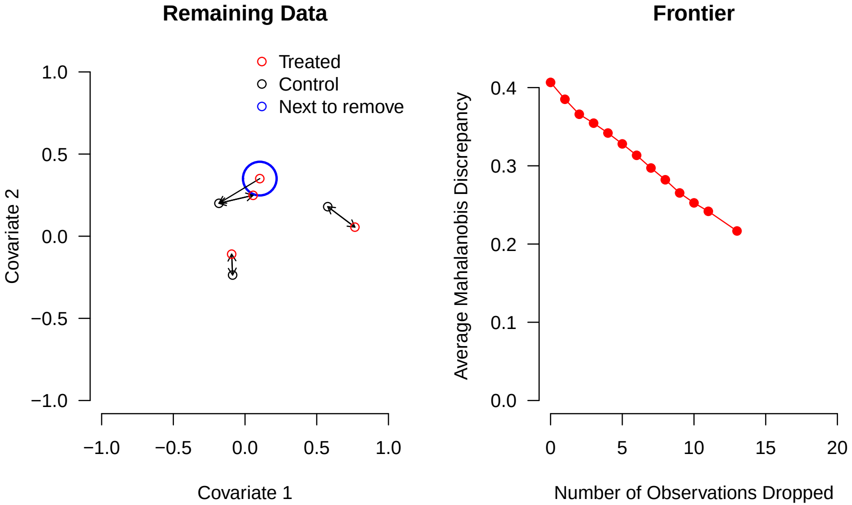
Constructing the Mahalanobis Frontier



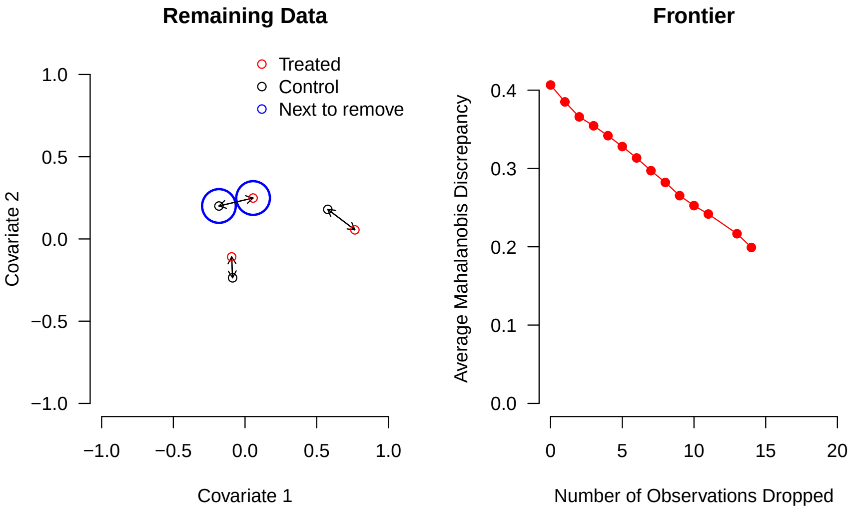
Constructing the Mahalanobis Frontier



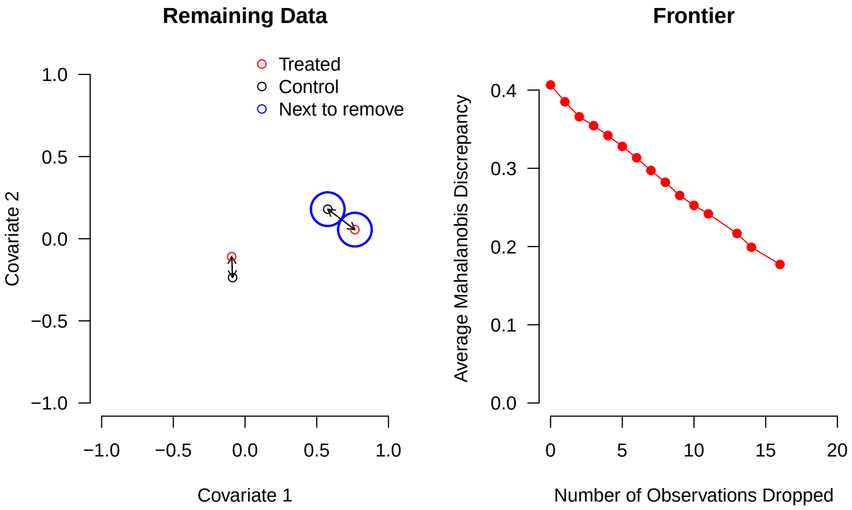
Constructing the Mahalanobis Frontier



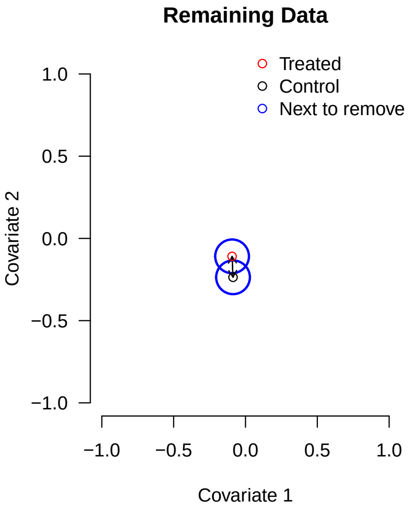
Constructing the Mahalanobis Frontier

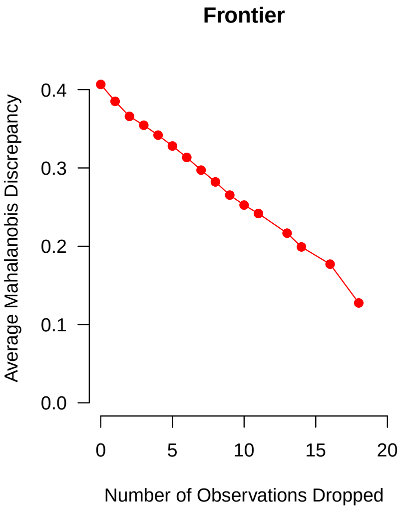


Constructing the Mahalanobis Frontier

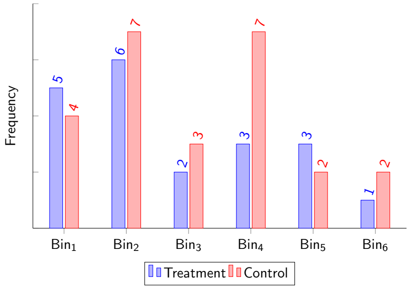


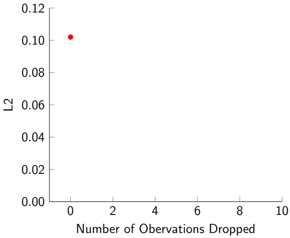
Constructing the Mahalanobis Frontier



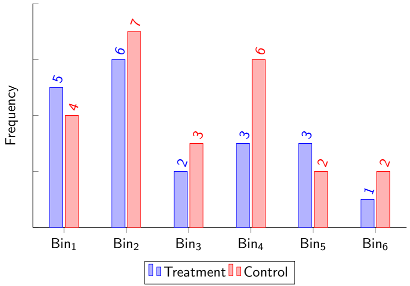


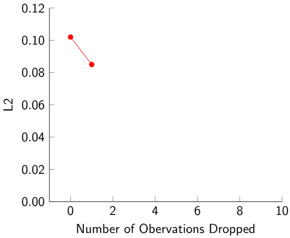
Constructing the L1/L2 Frontier



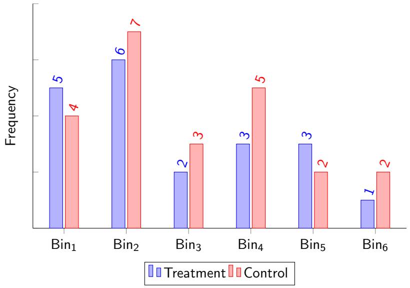


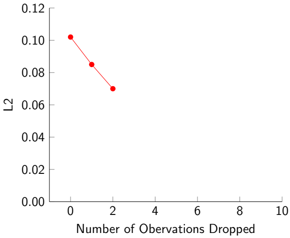
Constructing the L1/L2 Frontier



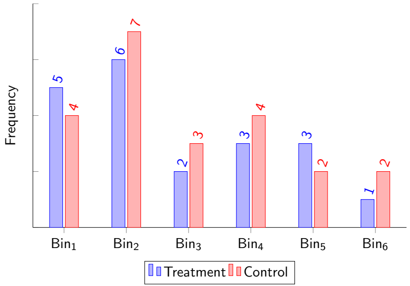


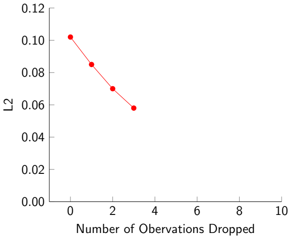
Constructing the L1/L2 Frontier



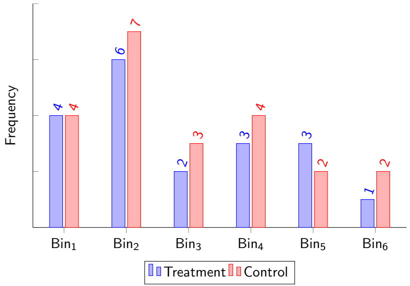


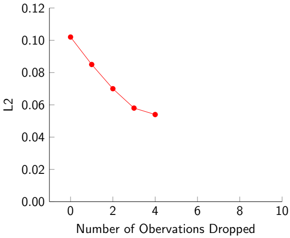
Constructing the L1/L2 Frontier



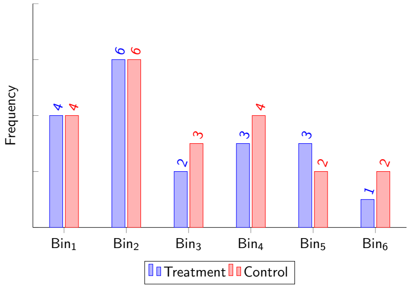


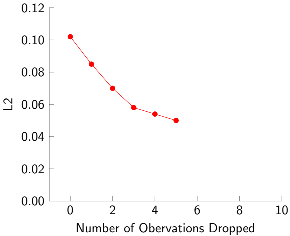
Constructing the L1/L2 Frontier



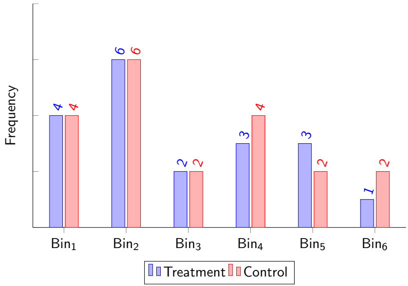


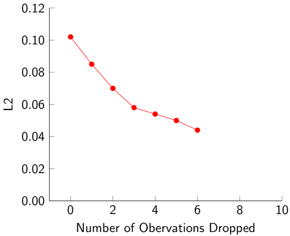
Constructing the L1/L2 Frontier



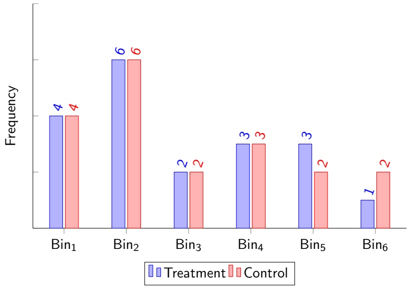


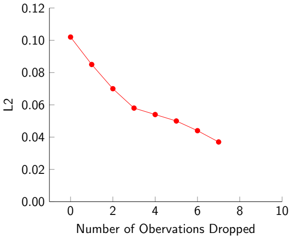
Constructing the L1/L2 Frontier



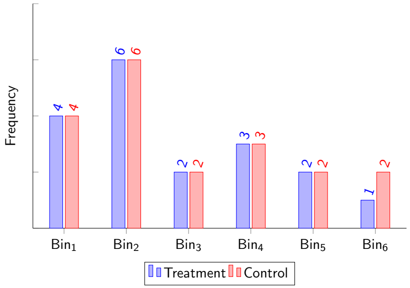


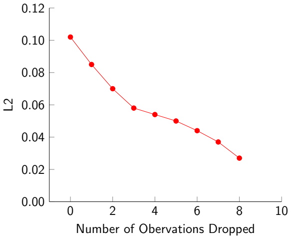
Constructing the L1/L2 Frontier



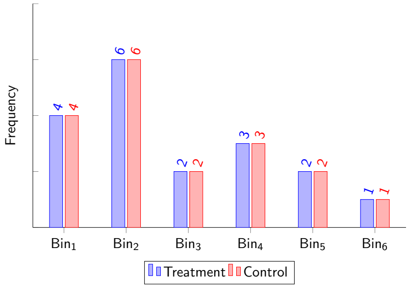


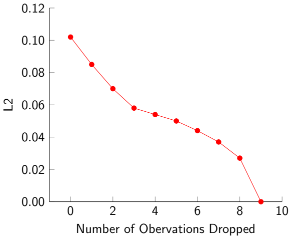
Constructing the L1/L2 Frontier

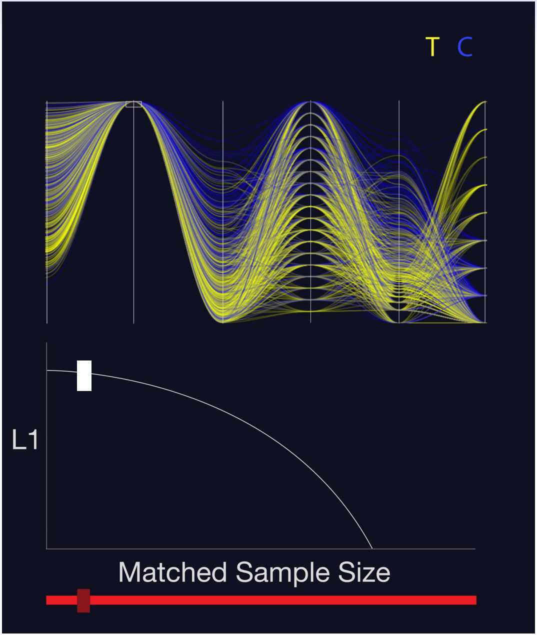




Constructing the L1/L2 Frontier

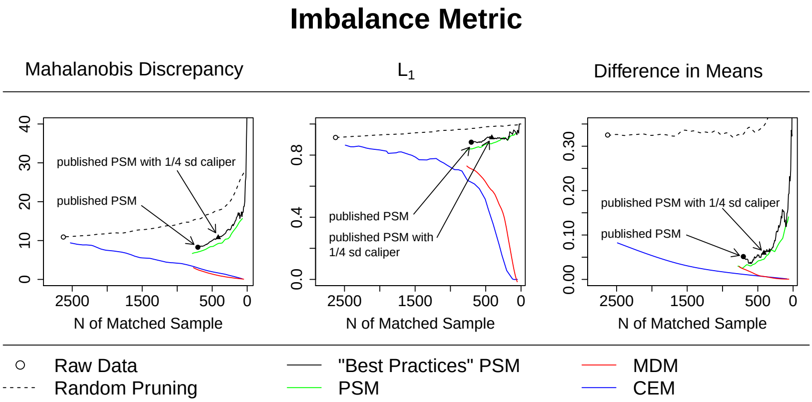





T C
L1
Matched Sample Size

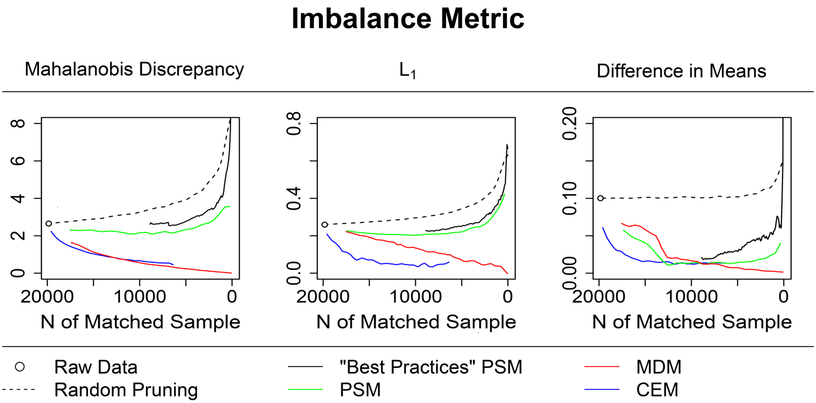
Foreign Aid Shocks & Conflict

King, Nielsen, Coberley, Pope, and Wells (2012)



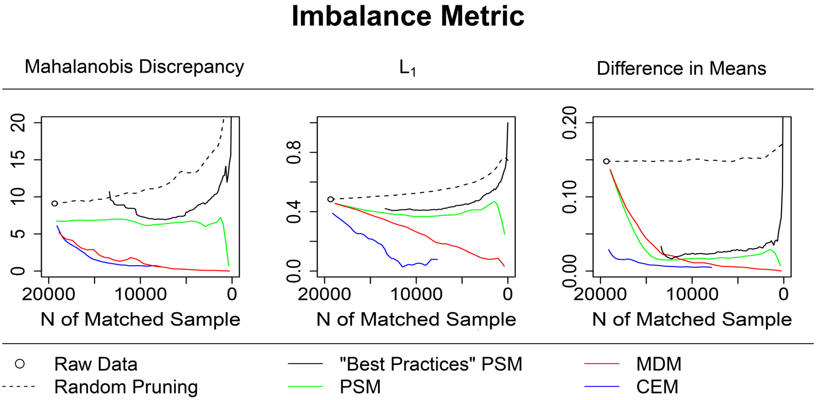
Healthways Data

King, Nielsen, Coberley, Pope, and Wells (2012)



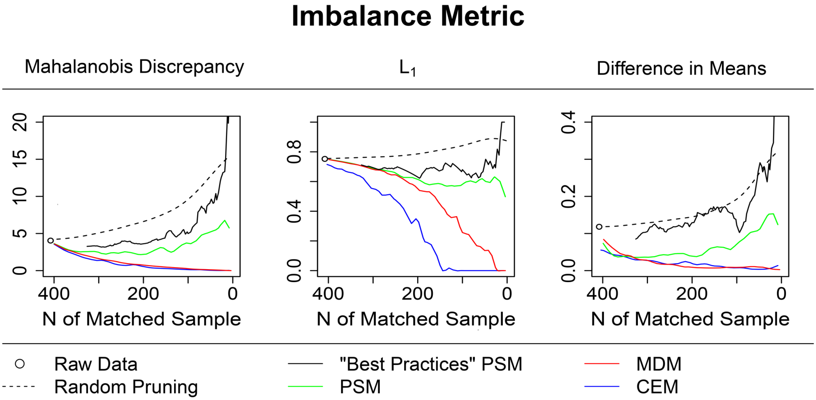
Called/Not Called Data

King, Nielsen, Coberley, Pope, and Wells (2012)



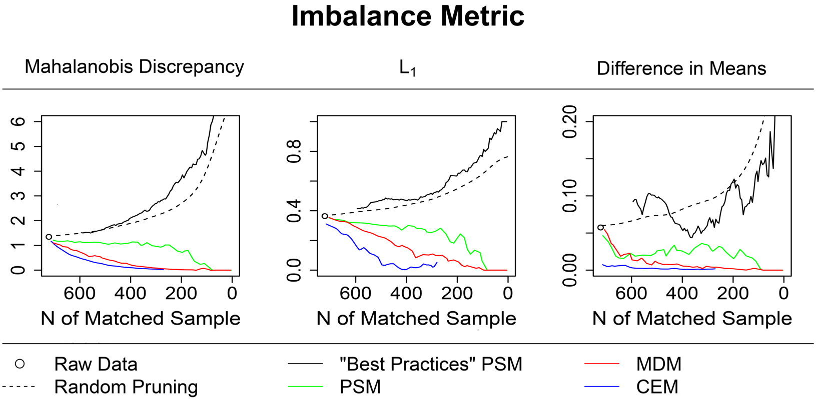
FDA Drug Approval Times

King, Nielsen, Coberley, Pope, and Wells (2012)

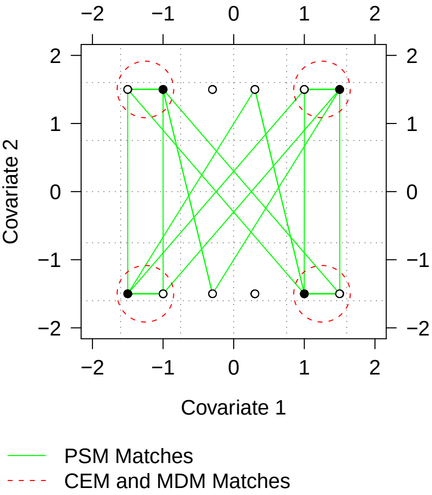


Job Training (Lelonde Data)

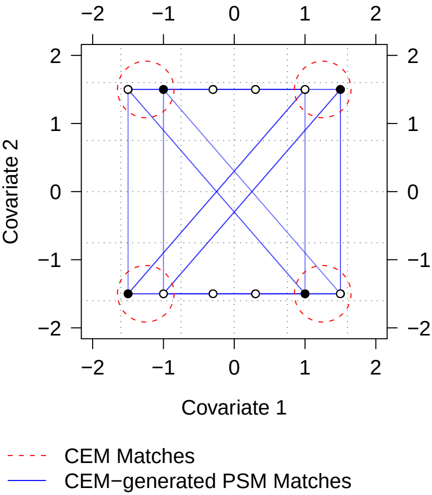
King, Nielsen, Coberley, Pope, and Wells (2012)



PSM Approximates Random Matching in Balanced Data



Destroying CEM with PSM's Two Step Approach



Conclusions

Conclusions

- The Matching Frontier

Conclusions

- The Matching Frontier
    - Easy to use; no need to iterate

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data
      (Random matching increases imbalance)

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake

Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake
    - 1/4 caliper on propensity score: mistake

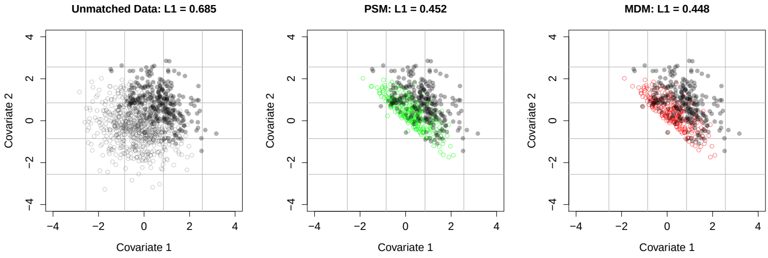
Conclusions

- The Matching Frontier
  - Easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake
    - 1/4 caliper on propensity score: mistake
- CEM and Mahalanobis do not have PSM's problems

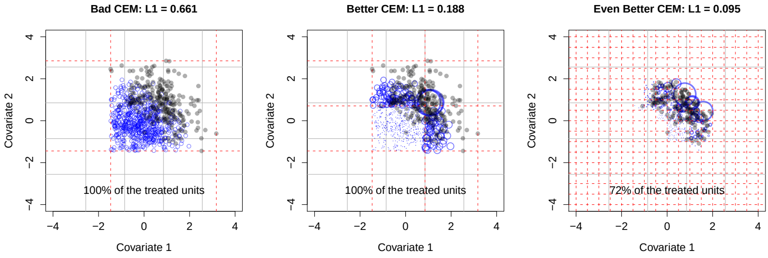
For more information,

GaryKing.org/cem

Data where PSM Works Reasonably Well — PSM & MDM



Data where PSM Works Reasonably Well — CEM



CEM Weights and Nonparametric Propensity Score

CEM Weight: $$w_i = \frac{m_i^T}{m_i^C} \quad (+\text{ normalization})$$

CEM Weights and Nonparametric Propensity Score

CEM Weight: $$w_i = \frac{m_i^T}{m_i^C} \quad (+\ \text{normalization})$$

CEM Pscore: $$\widehat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C}$$

CEM Weights and Nonparametric Propensity Score

CEM Weight: $$w_i = \frac{m_i^T}{m_i^C} \quad (+\text{ normalization})$$

CEM Pscore: $$\widehat{\Pr}(T_i = 1 | X_i) = \frac{m_i^T}{m_i^T + m_i^C}$$

$\rightsquigarrow$ CEM:

CEM Weights and Nonparametric Propensity Score

CEM Weight: $$w_i = \frac{m_i^T}{m_i^C} \quad (+\text{ normalization})$$

CEM Pscore: $$\widehat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C}$$

$\leadsto$ CEM:

- Gives a better pscore than PSM

CEM Weights and Nonparametric Propensity Score

CEM Weight: $$w_i = \frac{m_i^T}{m_i^C} \quad (+\text{ normalization})$$

CEM Pscore: $$\widehat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C}$$

$\rightsquigarrow$ CEM:

- Gives a better pscore than PSM
- Doesn't match based on crippled information